Eur Biophys J (2011) 40:843–856
DOI 10.1007/s00249-011-0700-9

ORIGINAL PAPER

Definition and testing of the GROMOS force-field versions 54A7 and 54B7

Nathan Schmid · Andreas P. Eichenberger · Alexandra Choutko · Sereina Riniker · Moritz Winger · Alan E. Mark · Wilfred F. van Gunsteren

Received: 18 January 2011 / Revised: 24 February 2011 / Accepted: 17 March 2011 / Published online: 30 April 2011

Abstract New parameter sets of the GROMOS biomolecular force field, 54A7 and 54B7, are introduced. These parameter sets summarise some previously published force field modifications: The 53A6 helical propensities are corrected through new φ/ψ torsional angle terms and a modification of the N–H, C=O repulsion, a new atom type for a charged $-CH_3$ in the choline moiety is added, the Na^+ and Cl^- ions are modified to reproduce the free energy of hydration, and additional improper torsional angle types for free energy calculations involving a chirality change are introduced. The new helical propensity modification is tested using the benchmark proteins hen egg-white lysozyme, fox1 RNA binding domain, chorismate mutase and the GCN4-p1 peptide. The stability of the proteins is improved in comparison with the 53A6 force field, and good agreement with a range of primary experimental data is obtained.

Keywords GROMOS · 54A7 · Force field · Secondary structure

Abbreviations

CM	Chorismate mutase
FOX	Fox1 RNA binding domain
GCN	GCN4-p1 peptide
HEWL	Hen egg-white lysozyme
PDB	Protein Data Bank
RMSD	Root-mean-square deviation
SPC	Simple point charge

Introduction

Biomolecular simulation involves four major challenges: (1) the force field must faithfully represent the atomic and molecular interactions, (2) the conformational space must be sampled in a manner which is both fast and efficient, (3) a Boltzmann configurational ensemble must be generated in order to reproduce thermodynamic quantities and (4) appropriate experimental data must be available against which the simulations can be validated. The quality of the force field is perhaps the most important of these issues and is the aspect addressed in this work. The interaction between the atoms in a system must be described with sufficient accuracy to reproduce the properties and mechanisms underlying the processes of interest. In the first generation of classical biomolecular force fields such as AMBER (Weiner and Kollman 1981; Pearlman et al. 1995; Cornell et al. 1995), CHARMM (Brooks et al. 1983; MacKerell et al. 1995, 1998), OPLS-AA (Jorgensen and Tirado-Rives 1988; Jorgensen et al. 1996) and GROMOS (van Gunsteren et al. 1996) the parameters were chosen so as to reproduce either spectroscopic or crystallographic structural data. Subsequently, increasing computer power

Electronic supplementary material The online version of this article (doi:10.1007/s00249-011-0700-9) contains supplementary material, which is available to authorized users.

N. Schmid · A. P. Eichenberger · A. Choutko · S. Riniker · W. F. van Gunsteren (✉)
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology ETH, 8093 Zürich, Switzerland
e-mail: wfvgn@igc.phys.chem.ethz.ch

M. Winger · A. E. Mark
School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

made the simulation of liquids possible and condensed-phase thermodynamic data such as densities, energies and free energies were included in the parameterisation, leading to second-generation force fields. For example, the GROMOS 45A4 (Daura et al. 1998; Schuler et al. 2001) parameter set was parameterised against the thermodynamic properties of aliphatic chains. In the subsequent generation of the GROMOS force field the polar amino acid side-chains and peptide backbone moiety were reparametrised. This resulted in the GROMOS 53A6 parameter set (Oostenbrink et al. 2004, 2005). However, the GROMOS 53A6 force field, in which the hydration properties of amino acid analogues are in good agreement with experiment, did not improve the stability of the dominant fold for all peptides (Oostenbrink et al. 2005). Rather, short α-helices were found to be less stable than expected. This suggested that the dihedral-angle parameters of the backbone transferred from the earlier version of the force field were no longer appropriate. Subsequently, a variety of torsional-angle potential energy functions were investigated by different workers. Cao et al. (2009) proposed a correction where the φ/ψ torsional angle terms were reparametrised and a torsional cross term depending on the sum of the φ- and ψ-angles was added. In an alternative approach leading to the 54A7 force field, the torsional angle terms were reparametrised based on fitting to a large set of high-resolution crystal structures (Xu et al. in preparation) and the N–O non-bonded interactions between the peptide nitrogen and oxygen atoms was adjusted to be less repulsive. In the present work we test these alternative approaches using four different test systems: hen egg-white lysozyme (HEWL, 129 residues), the fox1 RNA binding domain (FOX, 88 residues and 7 RNA bases), chorismate mutase (CM, 165 residues) and the GCN4-p1 peptide (GCN, 16 residues). Because the results for the two mentioned modifications yielded similar results, the simpler one was adopted in the new 54A7 parameter set. Recently, improved parameters for the simulation of lipids were reported (Poger et al. 2010). Incorporation in the GROMOS 53A6 force field required the definition of an additional atom type, which led to the 54 atom types of the 54A7 set. Finally, new Na^+ and Cl^- parameters from Reif and Hünenberger (2010) were added and an additional improper dihedral angle type was defined in order to facilitate free energy calculations involving a change in chirality. The definition of the changes from the 53A6 parameter set to the 54A7 parameter set is given in this work.

Definition of the GROMOS 54A7 force field

The GROMOS force field 54A7 is a modification of the GROMOS 53A6 force field, with four modifications:

1. The torsional-angle energy term for the polypeptide φ- and ψ-dihedral angles is modified in conjunction with a change of the combination prescription of the C_{12} van der Waals parameters for the atom type pair O(IAC = 1)–N(IAC = 6):
   (a) In the selection table for the repulsive van der Waals $C_{12}^{1/2}$(I,I) parameters, Table 8 of Oostenbrink et al. (2004), the type for the O(IAC = 1)–N(IAC = 6) pair (i.e. line 1, column 6) is changed from "2" to "1". This means that a $C_{12}^{1/2}$(O,O) value of 1.000×10^{-3} [kJ mol^{-1} nm^{12}]$^{1/2}$ for the O-atom (IAC = 1) is selected for the interaction with an N-atom (IAC = 6) compared with the $C_{12}^{1/2}$(O,O) value of 1.130×10^{-3} [kJ mol^{-1} nm^{12}]$^{1/2}$ in 53A6.
   (b) Four different dihedral torsional angle types are added to Table 5 of Oostenbrink et al. (2004):

Type code	$K_{\varphi n}$ [kJ mol^{-1}]	cos(δ_n)	m_n	Example	$K_{\varphi n}$ [kcal mol^{-1}]
42	3.50	−1	2	–CHn–C–	0.84
43	2.80	+1	3	–CHn–N–	0.67
44	0.70	−1	6	–CHn–N–	0.17
45	0.40	+1	6	–CHn–C–	0.10

In the molecular topology building blocks for the α-peptides and β-peptides the dihedral-angle type 39 (53A6) in the backbone C–N–CA–C dihedral angle (α-peptide) or the backbone C–N–CB–CA dihedral angle (β-peptide) is changed to type 44 (54A7) and the same dihedral angle with type 43 (54A7) is added. In addition, the dihedral-angle type 40 (53A6) for the backbone N–CA–C–N dihedral angle (α-peptide) or the backbone CB–CA–C–N dihedral angle (β-peptide) is to be changed to type 45 (54A7), and the same dihedral angle with type 42 (54A7) is added.

These changes increase the hydrogen-bonding interaction between the N–H and the C=O groups in the polypeptide backbone and bring the φ- and ψ-angle distributions for a number of proteins more in line with the preferences observed in PDB protein structures.

2. A new van der Waals non-bonded atom type for a charged -CH3 group (IAC = 54) is introduced in Table 6 of Oostenbrink et al. (2004) in order to increase the repulsion between the positively charged −CH3 groups of the choline moiety and the negatively charged −OM oxygen atoms of the phosphate moiety

in dipalmitoylphosphatidylcholine (DPPC)-type lipids, see Poger et al. (2010).

(a) In the normal van der Waals parameters table, Table 7 of Oostenbrink et al. (2004), an additional atom type is added:

IAC	Atom type	$[C6(I,I)]^{1/2}$ [kJ mol^{-1} nm^{6}]$^{1/2}$	$[C12(I,I)]^{1/2}$ 10^{-3} [kJ mol^{-1} nm^{12}]$^{1/2}$		
			1	2	3
54	CH3p	0.09805	5.162	–	–

(b) In the third-neighbour van der Waals parameters table, Table 9 of Oostenbrink et al. (2004), an additional atom type is added:

IAC	Atom type	$[C6(I,I)]^{1/2}$ [kJ mol^{-1} nm^{6}]$^{1/2}$	$[C12(I,I)]^{1/2}$ 10^{-3} [kJ mol^{-1} nm^{12}]$^{1/2}$		
			1	2	3
54	CH3p	0.08278	2.456	–	–

(c) In the selection table for the repulsive van der Waals $C_{12}^{1/2}(I,I)$ parameters, Table 8 of Oostenbrink et al. (2004), the matrix is enlarged by one column and one row to accommodate the new atom type. For all pairs, type 1 is selected with the exception of the OM(IAC = 2)-CH3p(IAC = 54) pair (line 2, column 54) where type 3 is selected.

These changes increase the area per lipid for DPPC bilayers, see Poger et al. (2010), by increasing the repulsion of their phosphate oxygen which leads to the Lennard–Jones parameters $\sigma = 0.3877$ nm and $\epsilon = 0.3433$ kJ mol^{-1} for the choline CH_3 (CH3p)-phosphate oxygen (OM) pair.

3. The van der Waals non-bonded interaction parameters for the Na^+ and Cl^- ions are taken from the set L proposed by Reif and Hünenberger (2010).

(a) In the normal van der Waals parameters table, Table 7 of Oostenbrink et al. (2004), the parameters for Na^+ (IAC = 37) and Cl^- (IAC = 38) are set to:

IAC	Atom type	$[C6(I,I)]^{1/2}$ [kJ mol^{-1} nm^{6}]$^{1/2}$	$[C12(I,I)]^{1/2}$ 10^{-3} [kJ mol^{-1} nm^{12}]$^{1/2}$		
			1	2	3
37	NA^+	0.0088792	0.2700	0.2700	–
38	CL^-	0.11318	7.776	7.776	7.776

The third-neighbour CS6 and CS12 parameters were also changed accordingly.

These changes bring the solvation properties of Na^+ and Cl^- more in line with experiment.

4. To facilitate the calculation of differences in free energy involving changes in chirality, two additional improper dihedral-angle types are defined, i.e. one with −35° and one with 180° as energy minimum.

(a) The two new improper (harmonic) dihedral-angle types are added to Table 4 of Oostenbrink et al. (2004):

Type code	$K_{\xi n}$ [kJ mol^{-1} deg^{-2}]	ξ_{0n} [deg]	Example of usage	$K_{\xi n}$ [kcal mol^{-1} rad^{-2}]
4	0.0510	180.0	Planar groups	40
5	0.102	−35.26439	Tetrahedral centres	80

Corresponding changes in the 53B6 force field for in vacuo simulations yield the 54B7 force field for in vacuo simulations.

Materials and methods

All simulations were performed using a modified version of the GROMOS biomolecular simulation software (Christen et al. 2005) in conjunction with the parameter set of the GROMOS force field indicated: 45A4 (Schuler et al. 2001), 53A6 (Oostenbrink et al. 2004), Liu (Cao et al. 2009) or 54A7. Note that the CM system was not simulated using the Liu modification. Also, the GCN system was only simulated using the 53A6 and 54A7 force fields, meaning that a longer timescale could be examined. The dihedral-angle potential energy function in the GROMOS force field for the backbone φ- and ψ-angles is defined as

Table 1 Force-field parameters of the dihedral-angle term for peptide backbone φ- and ψ-torsional angles in the GROMOS force field Oostenbrink et al. (2004) and from Cao et al. (2009)

Term	45A4 and 53A6 (Oostenbrink et al. 2004)			54A7			Liu (Cao et al. 2009)		
	K/kJ mol^{-1}	m	δ	K/kJ mol^{-1}	m	δ	K/kJ mol^{-1}	m	δ
φ	1.0	6	180°	2.8	3	0°	1.2	1	0°
				0.7	6	180°	0.5	1	−120°
							0.8	2	180°
							1.0	3	0°
ψ	1.0	6	0°	3.5	2	180°	0.8	1	150°
				0.4	6	0°	0.5	3	−90°
							0.5	4	90°
$\varphi + \psi$							0.5	1	90°
							1.1	2	0°
							0.5	3	−135°
							0.5	4	120°
Non-bonded	C_{12}(N,O) type 2			C_{12}(N,O) type 1			C_{12}(N,O) type 2		

$$V^{\mathrm{GROMOS}}(\varphi,\psi) = \sum_{i=1}^{N_\varphi} K_i[1+\cos(m_i\varphi-\delta_i)] + \sum_{i=1}^{N_\psi} K_i[1+\cos(m_i\psi-\delta_i)], \quad (1)$$

where N_φ and N_ψ are the number of terms for one dihedral angle, see Table 1. In the approach of Liu and coworkers, a cross term that depends on the sum of the φ- and ψ-angle,

$$V^{\mathrm{cross}}(\varphi,\psi) = \sum_{i=1}^{N_{\varphi,\psi}} K_i[1+\cos(m_i(\varphi+\psi)-\delta_i)] \quad (2)$$

is added to the potential energy function (Eq. 1), resulting in the complete potential energy function for the backbone φ,ψ dihedral angles. The parameters of the different force fields are summarised in Table 1. The 45A4 and 53A6 force fields use the same description of the torsional potential energy term. In the 54A7 force field, these terms are adjusted and the repulsive (C_{12}) term of the Lennard–Jones potential energy term is changed from type 2 to type 1, which means that it is less repulsive.

The initial coordinates of the protein and RNA molecules were taken from structures (Fig. 1) deposited in the Protein Data Bank (PDB). The entry codes were 1AKI for HEWL (Artymiuk et al. 1982), 2ERR for FOX (Auweter et al. 2006), 2FP2 for CM (Okvist et al. 2006) and 2OVN for GCN (Steinmetz et al. 2007). In the case of 2ERR and 2OVN, the first structure of the NMR bundle was taken. In the case of 2FP2, only the first subunit of the dimeric protein was taken. All hydrogens were (re)generated by the GROMOS++ (Christen et al. 2005) program *gch*. Each system was first energy minimised in vacuo, then the protein plus RNA molecules in the case of FOX were solvated in cubic boxes filled with simple point charge (SPC) water (Berendsen et al. 1981) molecules. Periodic

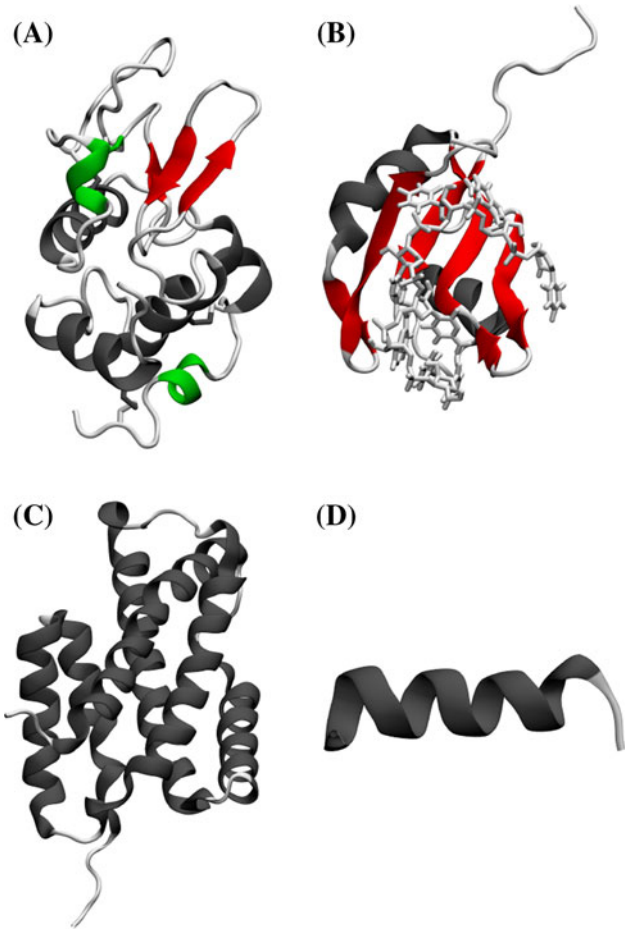

Fig. 1 Ribbon pictures of the four proteins investigated: **a** HEWL, α-helices (residues 5–14, 25–34, 89–100 and 109–114), 3_{10}-helices (residues 80–83 and 120–123) and β-sheet (residues 43–45, 51–53 and 58–59). **b** FOX, α-helices (residues 21–32 and 58–69) and β-sheet (residues 10–14, 38–40, 50–54, 72–73, 76–77 and 79–82). The nucleic acid is shown in ball-and-stick representation. **c** CM, α-helices (residues 4–17, 18–28, 35–52, 56–84, 85–89, 95–116, 117–121, 124–141 and 144–155). **d** GCN, α-helix (residues 1–16). Colour code: α-helix (*black*), 3_{10}-helix (*green*) and β-sheet (*red*)

boundary conditions were applied. The boxes contained 14,378 (HEWL using the 45A4 force field), 14,355 (HEWL using the other force fields), 11,533 (FOX, 45A4), 11,552 (FOX using the other force fields), 24,040 (CM, 45A4), 23,959 (CM using the other force fields) or 2,894 (GCN) solvent molecules. Each system was energy minimised to relax the solvent with the positions of the heavy atoms within the protein and RNA restrained. All simulations were initiated using the following equilibration scheme: First, the initial velocities were randomly generated from a Maxwell–Boltzmann distribution at 60 K. All solute atom positions were restrained to their positions in the initial structure through a harmonic potential energy term with force constant of 2.5×10^4 kJ mol^{-1} nm^{-2}. The system was simulated with these settings for 20 ps. Before each of three consecutive 20 ps simulations, the temperature was raised by 60 K with the positional restraints being reduced by a factor of 10 at each step. Before the next 60 ps simulation the position restraints were removed and the temperature was raised to 300 K, resulting in the starting configurations for the main sampling runs. Next, 20 or 50 ns production simulations were performed. The temperature of 300 K and atmospheric pressure were kept constant using a weak-coupling approach (Berendsen et al. 1984) with relaxation times $\tau_T = 0.1$ ps and $\tau_p = 0.5$ ps and isothermal compressibility of 4.575×10^{-4} (kJ mol^{-1} nm^{-3})$^{-1}$. Non-bonded interactions were calculated using a triple-range cutoff scheme. The interactions within a cutoff distance of 0.8 nm were calculated at every step from a

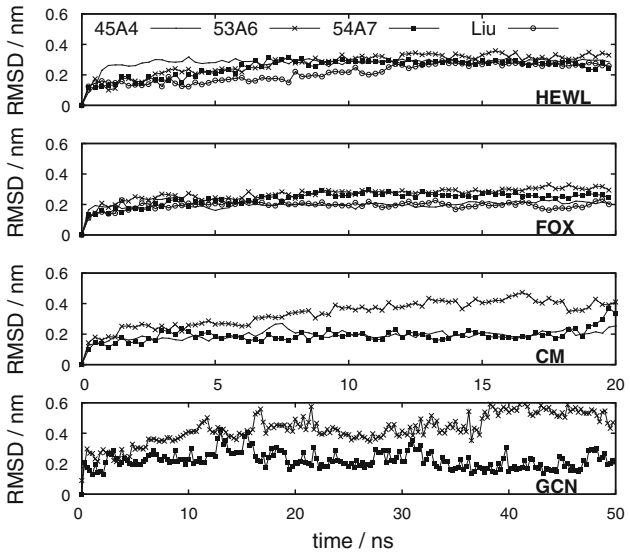

Fig. 2 Root-mean-square deviations of backbone atoms of the four protein structures from their starting structure. For HEWL and FOX four, for CM three and for GCN two different force fields were used

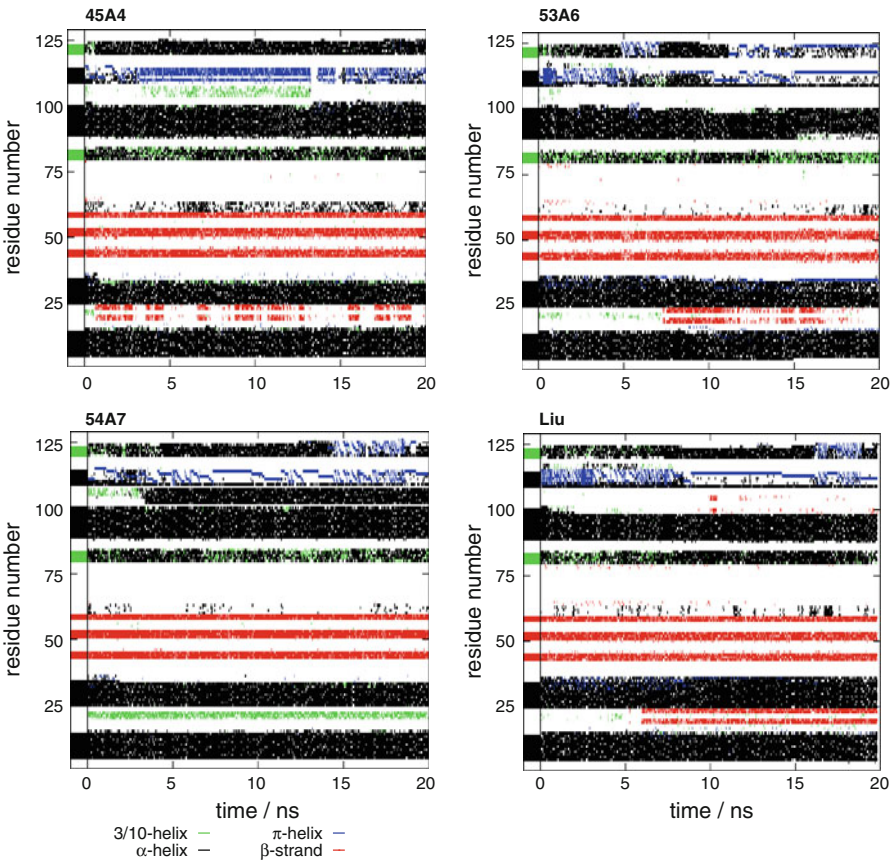

Fig. 3 Secondary structure of HEWL in MD simulations using four different force fields. The definition of Kabsch and Sander (1983) is used. The secondary structure suggested in the PDB is shown in the box at the residue axis. Colour code: α-helix (*black*), 3_{10}-helix (*green*), π-helix (*blue*) and β-sheet (*red*)

pair list which was updated every fifth time step. At this point, interactions between atoms (of charge groups) within 1.4 nm were also calculated and were kept constant between updates. To account for the influence of the dielectric medium outside the cutoff sphere of 1.4 nm, a reaction-field force based on a relative dielectric permittivity ϵ of 61 (Heinz et al. 2001) was added. Bond lengths and the bond angle of the water molecules were constrained using the SHAKE algorithm (Ryckaert et al. 1977). Centre-of-mass motion of the whole system was removed every 2 ps. System configurations were saved every 0.2 ps for analysis.

Analysis

The atom-positional root-mean-square deviation (RMSD) between the backbone atoms of two structures was calculated after the superposition of these atoms. The secondary structure of the protein was assigned according to the DSSP criteria defined by Kabsch and Sander (1983). The presence of a hydrogen bond was assigned based on the following geometric criteria. If the hydrogen-acceptor distance was less than 0.25 nm and the donor–hydrogen–acceptor angle was at least 135°, the hydrogen bond was considered to be present. Proton–proton distances were compared with upper bounds derived from NMR spectra (Auweter et al. 2006; Schwalbe et al. 2001). Proton–proton distances were averaged using r^{-6} averaging of $\overline{r} = \langle r^{-6} \rangle^{-1/6}$. Positions of protons that were not treated explicitly by the force field were calculated from standard configurations (van Gunsteren et al. 1996). In cases where the NOE upper bounds corresponded to more than one proton, a pseudo-atom approach (Wüthrich et al. 1983) with the standard GROMOS corrections (van Gunsteren et al. 1996; Oostenbrink et al. 2005) was applied. No additional multiplicity corrections (Fletcher et al. 1996) were added. ^3J-coupling constants $^3J(\mathrm{H_N} - \mathrm{H_\alpha})$ were calculated using the Karplus (1959) relation

$$^3J_{\mathrm{H_N}-\mathrm{H_\alpha}} = A\cos\varphi'^2 + B\cos\varphi' + C, \quad (3)$$

where φ' is the dihedral angle between the planes defined by the atoms (H, N, C_α) and the atoms (N, C_α, H_α). $^3J_{\mathrm{HN}} - \mathrm{H_\alpha}$ was calculated as a function $^3J_{\mathrm{H_N}-\mathrm{H_\alpha}}(\varphi)$ of the backbone angle $\varphi = (\mathrm{C} - \mathrm{N} - \mathrm{C_\alpha} - \mathrm{C})$, with $\varphi = \varphi' + 60°$. The parameters A, B and C were set to 6.4, −1.4, and 1.9 Hz, respectively (Pardi et al. 1984).

Software and hardware

All simulation and energy-minimisation computations were carried out using MD++ 1.0 of the GROMOS05 package (Christen et al. 2005). For analysis, GROMOS++ 1.0 (Christen et al. 2005) was used. Additional analysis, conversion and batch programs were written in the Perl programming language. Visualisation was done with Visual Molecular Dynamics (VMD) software (Humphrey et al. 1996).

Table 2 Occurrence (%) of backbone hydrogen bonds of HEWL for the elements of secondary structure proposed in the PDB file 1AKI

Donor		Acceptor		45A4	53A6	54A7	Liu
α-helix 1							
9	Ala	5	Arg	89	82	97	94
10	Ala	6	Cys	95	95	96	97
11	Ala	7	Glu	89	90	83	89
12	Met	8	Leu	98	91	98	97
13	Lys	9	Ala	94	89	97	98
14	Arg	10	Ala	82	94	91	95
α-helix 2							
29	Val	25	Leu	89	84	88	73
30	Cys	26	Gly	96	98	99	97
31	Ala	27	Asn	94	96	97	96
32	Ala	28	Trp	88	85	71	95
33	Lys	29	Val	38	87	59	91
34	Phe	30	Cys	6	87	63	95
β-sheet							
44	Asn	52	Asp	72	90	74	82
52	Asp	44	Asn	84	91	89	89
53	Tyr	58	Ile	82	96	94	96
58	Ile	53	Tyr	93	92	94	98
3_{10}-helix 1							
83	Leu	80	Cys	41	56	50	43
α-helix 3							
93	Asn	89	Thr	94	49	95	93
94	Cys	90	Ala	92	79	97	96
95	Ala	91	Ser	98	99	97	97
96	Lys	92	Val	93	81	77	76
97	Lys	93	Asn	54	63	25	33
98	Ile	94	Cys	89	89	78	85
99	Val	95	Ala	76	44	88	31
100	Ser	96	Lys	59	56	70	7
α-helix 4							
113	Asn	109	Val	14	74	65	79
114	Arg	110	Ala	8	23	15	19
3_{10}-helix 2							
123	Trp	120	Val	15	8	8	6

Results and discussion

The HEWL and FOX systems were simulated for 20 ns using four different parameter sets: those of the GROMOS

force fields 45A4, 53A6 and 54A7 and that of Liu. The CM system was also simulated for 20 ns but using only three different parameter sets: 45A4, 53A6 and 54A7. The GCN system was simulated for 50 ns but only using the 53A6 and 54A7 parameter sets. Time series of the backbone atom-positional root-mean-square deviations with respect to the initial structure were calculated from the trajectories generated during the simulations (Fig. 2). For the HEWL

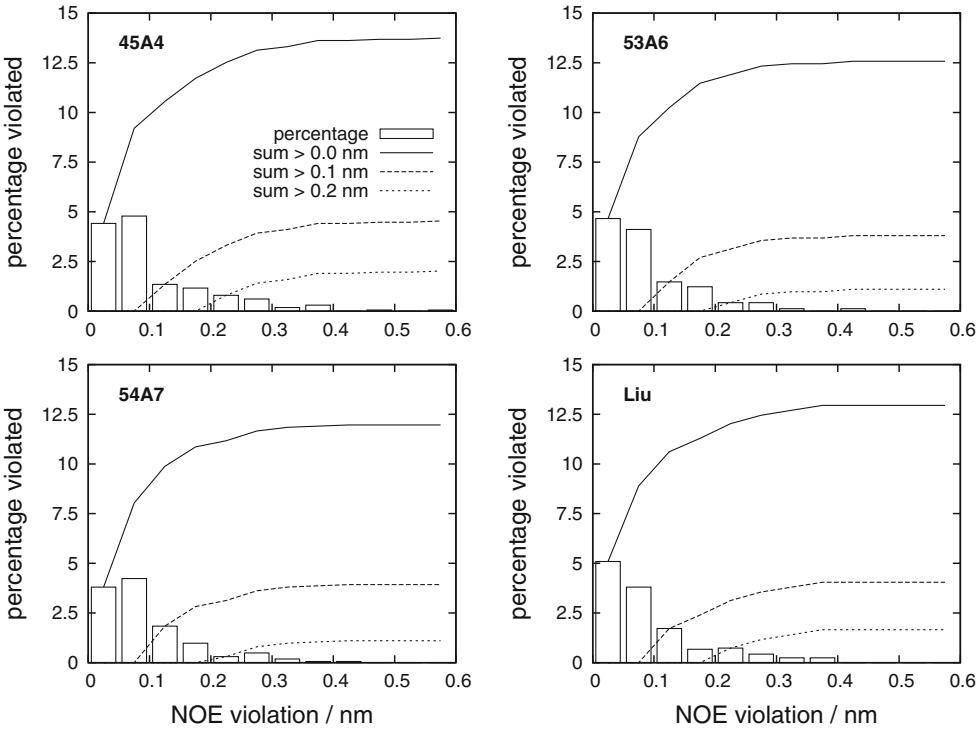

Fig. 4 NOE upper bound violations of HEWL in MD simulations using four different force fields. Total number of NOE bounds: 1,630

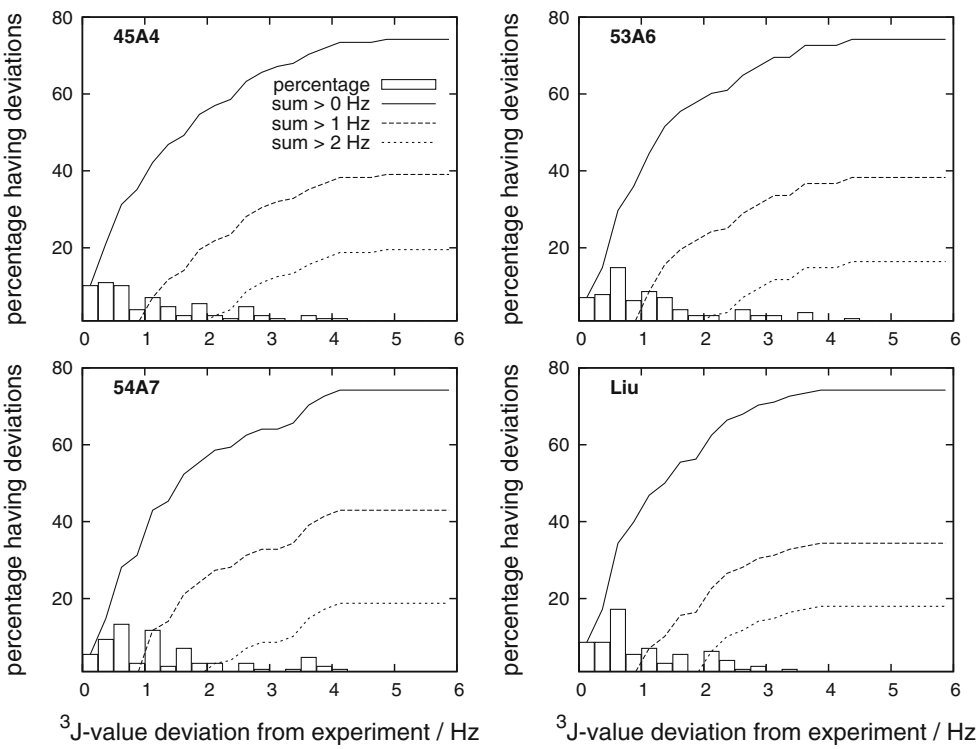

Fig. 5 ^3J-coupling deviations from measured values of HEWL in MD simulations using four different force fields. Total number of ^3J-couplings: 128

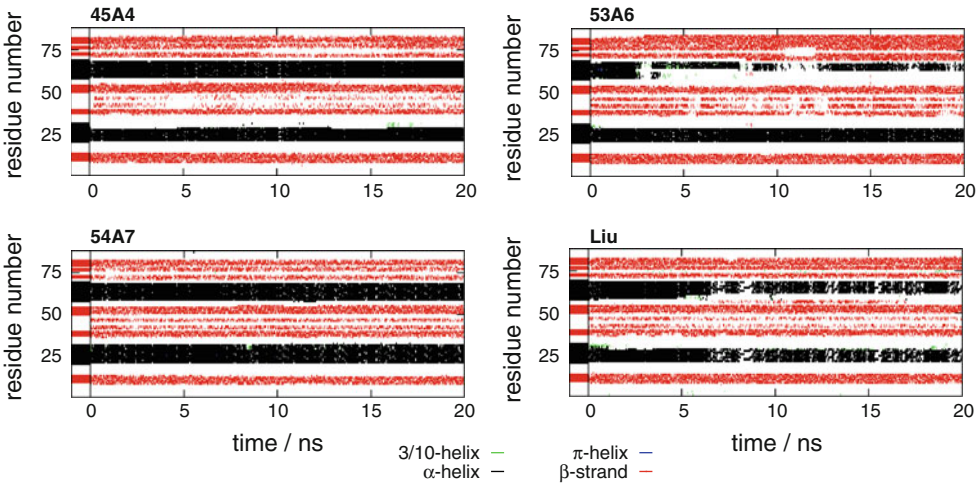

Fig. 6 Secondary structure of FOX in MD simulations using four different force fields. The definition of Kabsch and Sander (1983) is used. The secondary structure suggested in the PDB is shown in the box at the residue axis. Colour code: α-helix (*black*), 3_{10}-helix (*green*), π-helix (*blue*) and β-sheet (*red*)

Table 3 Occurrence (%) of backbone hydrogen bonds of FOX for the elements of secondary structure proposed in the PDB file 2ERR

Donor		Acceptor		45A4	53A6	54A7	Liu
β-sheet							
9	Lys	56	Glu	47	13	37	85
11	Leu	53	Val	96	57	93	96
13	Val	51	Gly	94	96	90	95
14	Ser	79	Glu	75	73	65	67
37	Asp	54	Thr	0	10	0	1
39	Glu	52	Phe	86	87	95	91
41	Ile	50	Phe	45	6	95	94
43	Asn	46	Gly	79	89	96	97
46	Gly	43	Asn	14	57	44	25
48	Lys	41	Ile	69	94	99	99
51	Gly	13	Val	57	95	44	91
52	Phe	39	Glu	87	94	92	94
53	Val	11	Leu	90	97	98	97
54	Thr	37	Asp	77	68	92	79
55	Phe	9	Lys	79	10	77	59
79	Glu	14	Ser	7	29	9	77
α-helix 1							
25	Leu	21	Arg	95	98	98	96
26	Arg	22	Asp	87	82	93	92
27	Gln	23	Pro	92	97	95	96
28	Met	24	Asp	81	96	96	89
29	Phe	25	Leu	74	89	98	70
30	Gly	26	Arg	39	89	87	43
31	Gln	27	Gln	0	0	59	0
32	Phe	28	Met	0	0	19	0
α-helix 2							
62	Asp	58	Ser	90	23	89	20
63	Arg	59	Ala	80	8	88	23
64	Ala	60	Asp	90	11	95	32
65	Arg	61	Ala	81	11	63	94
66	Glu	62	Asp	50	9	45	93
67	Lys	63	Arg	49	30	83	91
68	Leu	64	Ala	92	66	96	76
69	His	65	Arg	69	32	71	96

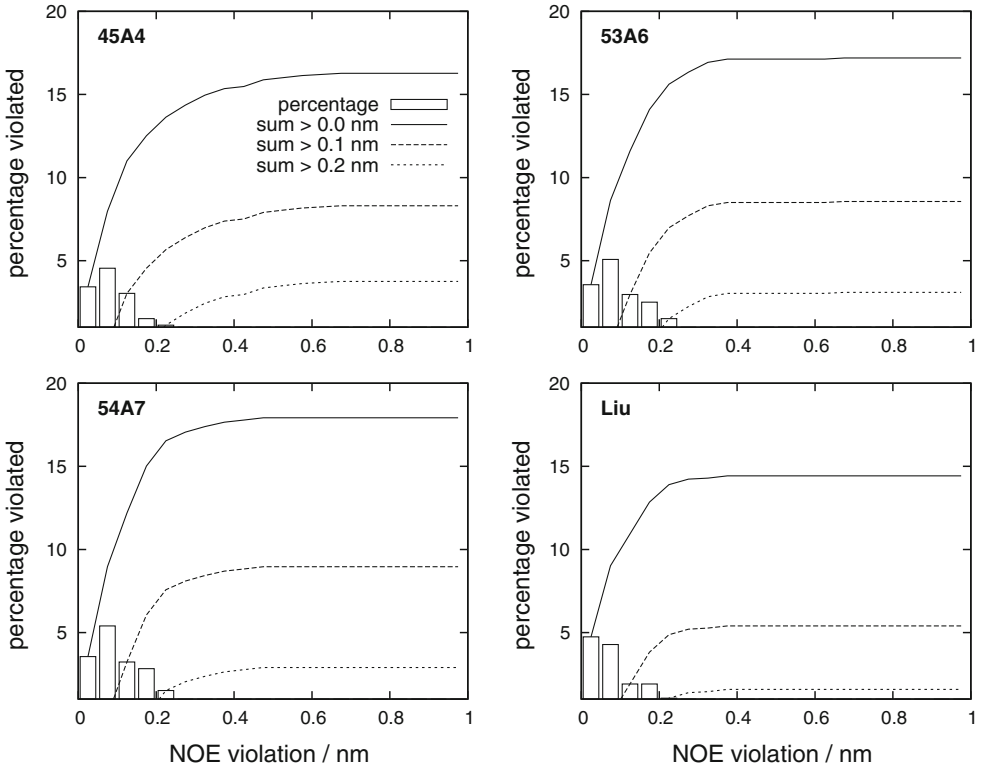

Fig. 7 NOE upper bound violations of FOX in MD simulations using four different force fields. Total number of NOE bounds: 1,518

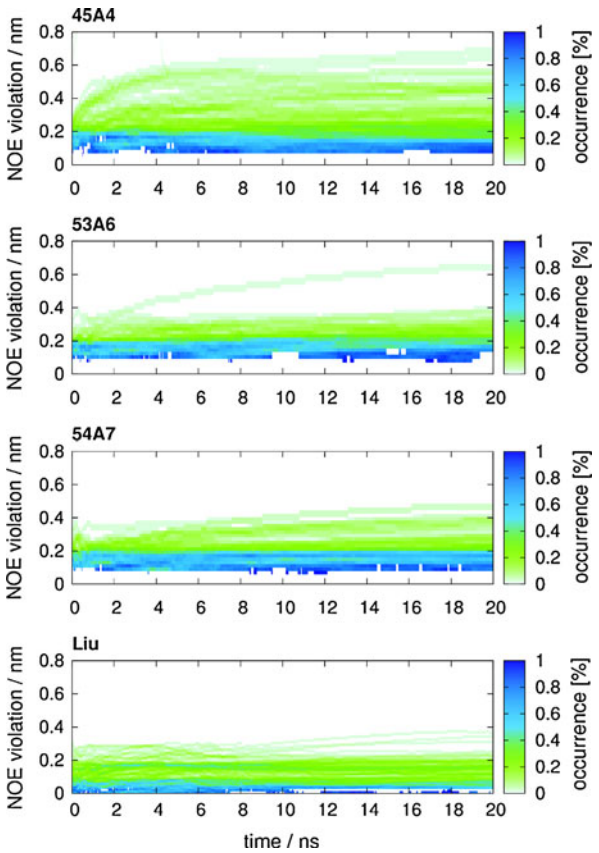

Fig. 8 NOE bound violation time series of FOX in MD simulations using four different force fields

system (top panel), the RMSD after 20 ns was similar for the four different force fields used. The 53A6 parameter set resulted in the highest RMSD values $\approx$0.35 nm, while 45A4 and 54A7 yielded slightly smaller values of $\approx$0.3 nm. Initially, the RMSD in the case of the Liu force field remained low, but after 12 ns there was no significant difference in comparison with the 54A7 force field. In the case of the FOX protein nucleic-acid complex, the RMSD was dominated by fluctuations in the tail and loop regions. For this reason only the backbone atoms of residues 9–82 were used in the calculation of the RMSD. The initial conformation was essentially conserved in all the simulations. The 53A6 force field again resulted in the largest RMSD value ($\approx$0.3 nm). The 45A4 and Liu force fields yielded values of $\approx$0.2 nm. The value for the 54A7 force-field parameter set was $\approx$0.24 nm, which is within the fluctuations in the system. The CM protein was significantly less stable using the 53A6 force field than using either the 45A4 or 54A7 force fields. The RMSD increase seen at the end of the 54A7 simulation will be explained below. The GCN system showed RMSD values of up to 0.6 nm in the 53A6 simulation, which is large given the small size of the system. The 54A7 force field improved the maintenance of the initial structure significantly but did not prevent unfolding and refolding events.

The main elements of secondary structure (Fig. 1), the large α-helices and β-sheets, of HEWL were conserved in all simulations (Fig. 3; Table 2). However, the small

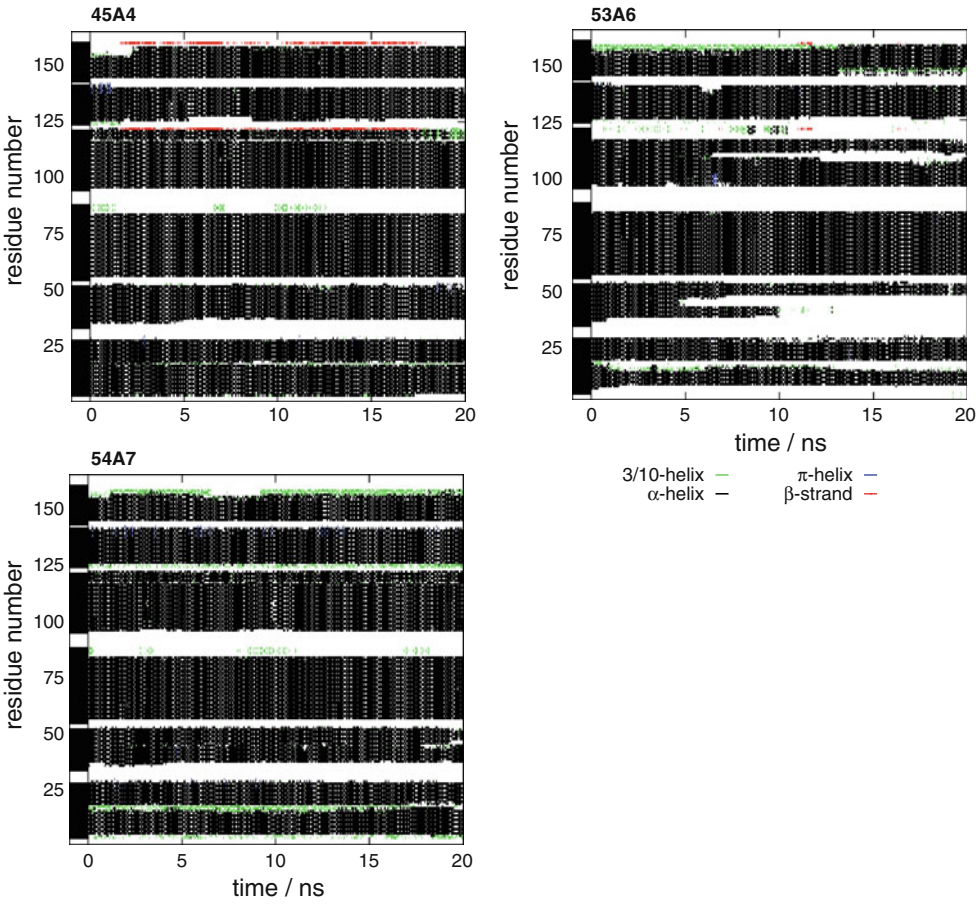

Fig. 9 Secondary structure of CM in MD simulations using three different force fields. The definition of Kabsch and Sander (1983) is used. The secondary structure suggested in the PDB is shown in the box at the residue axis. Colour code: α-helix (*black*), 3_{10}-helix (*green*), π-helix (*blue*) and β-sheet (*red*)

elements of secondary structure (α-helix 4 and 3_{10}-helix 1 and 2) were only marginally stable. There was a tendency for the sections of 3_{10}-helix to convert to α-helix. α-Helix 4 had a tendency to form a π-helix in all the simulations. The 45A4, 53A6 and Liu force fields had a tendency to stabilise β-sheet with regions of coil converted to β-sheet (i.e. residues 19, 20, 23 and 24) . The 54A7 force field in contrast stabilised the larger α-helices. The Liu force field failed to maintain the short 3_{10}-helices. Overall, the 54A7 force field conserved the secondary structure best. This said, the fact that some of the 3_{10}-helical structures were converted to α-helices hints at slight overstabilisation of the elements of α-helix. We note, however, that the ability to infer errors in the force field based on comparison of simulated structures versus those derived from NMR or X-ray data is limited, because the latter are themselves models into which force-field data has been incorporated during the process of structure determination (Glättli and van Gunsteren 2004, Dolenc et al. 2010). To validate the force fields, direct comparisons with observed experimental data are more valuable. The NOE bound violations (Fig. 4) are low in all simulations. The 45A4 and Liu force fields gave rise to the largest number of violations, but the differences from the 53A6 and 54A7 force fields were marginal. The integrated distributions give insight into the overall number of violations. The 53A6 and 54A7 force fields resulted in significantly lower integrated values. These force fields had approximately half the number of violations >0.2 nm in comparison with the 45A4 and Liu force fields. In contrast, the ^3J-couplings (Fig. 5) appear to be insensitive to the differences between the force-field parameter sets. We note, however, that the calculated ^3J-couplings have an uncertainty of at least 1 Hz due to the approximative, empirical character of Eq. 3 and the relation between structure and ^3J-coupling, and experimentally represent averages over much longer timescales than accessed during these simulations.

The FOX protein consists of two α-helices packed against a β-sheet (Fig. 1). The helices are exposed to solvent, which makes them less stable then those found in HEWL. In fact, the helices did not remain stable using the 45A4, 53A6 and Liu force fields (Fig. 6; Table 3). The first α-helix became shorter with loss of two hydrogen bonds. The second α-helix was not stable in the simulations using the 53A6 and Liu force fields. In particular, using the 53A6 force field, the helix was almost completely lost by the end of the simulation. In contrast, both helices remained stable when using the 54A7 force field. The 54A7 and the Liu

Table 4 Occurrence (%) of backbone hydrogen bonds of CM for the elements of secondary structure proposed in the PDB file 2FP2

Donor		Acceptor		45A4	53A6	54A7
α-helix 1						
8	Leu	4	Gln	33	0	4
9	Val	5	Leu	75	1	58
10	Asp	6	Ala	96	5	98
11	Ala	7	Glu	90	96	93
12	Ala	8	Leu	95	61	95
13	Ala	9	Val	97	97	97
14	Glu	10	Asp	89	84	91
15	Arg	11	Ala	91	25	95
16	Leu	12	Ala	90	10	65
17	Glu	13	Ala	80	0	22
α-helix 2						
22	Val	18	Val	45	52	62
23	Ala	19	Ala	79	97	84
24	Ala	20	Asp	94	92	97
25	Phe	21	Pro	85	90	96
26	Lys	22	Val	91	98	96
27	Trp	23	Ala	90	96	96
28	Arg	24	Ala	89	40	97
α-helix 3						
39	Val	35	Asp	2	1	0
40	Glu	36	Ser	17	2	10
41	Gln	37	Gly	77	36	74
42	Gln	38	Arg	90	29	91
43	Leu	39	Val	95	47	96
44	Ala	40	Glu	86	19	91
45	Lys	41	Gln	88	17	74
46	Leu	42	Gln	93	18	60
47	Gly	43	Leu	81	16	24
48	Glu	44	Ala	71	15	20
49	Asp	45	Lys	78	19	43
50	Ala	46	Leu	89	29	78
51	Arg	47	Gly	73	86	86
52	Ser	48	Glu	63	86	83
α-helix 4						
60	Val	56	Asp	81	79	83
61	Thr	57	Pro	93	94	93
62	Arg	58	Asp	78	78	86
63	Val	59	Tyr	81	90	96
64	Phe	60	Val	92	91	93
65	Asp	61	Thr	72	84	72
66	Asp	62	Arg	25	69	63
67	Gln	63	Val	27	48	51
68	Ile	64	Phe	96	96	95
69	Arg	65	Asp	95	95	96
70	Ala	66	Asp	90	87	92
71	Thr	67	Gln	83	83	85
72	Glu	68	Ile	84	91	93
73	Ala	69	Arg	73	74	71
74	Ile	70	Ala	73	77	62
75	Glu	71	Thr	86	76	94
76	Tyr	72	Glu	91	84	93
77	Ser	73	Ala	91	83	79
78	Arg	74	Ile	93	97	96
79	Phe	75	Glu	86	96	94
80	Ser	76	Tyr	87	90	93
81	Asp	77	Ser	76	81	73
82	Trp	78	Arg	85	89	91
83	Lys	79	Phe	80	71	86
84	Leu	80	Ser	82	79	83
α-helix 5						
89	Ala	85	Asn	3	0	1
α-helix 6						
99	Ser	95	Asp	81	83	68
100	Arg	96	Leu	96	82	77
101	Ser	97	Ser	93	91	89
102	Ala	98	Ala	83	78	86
103	Ile	99	Ser	79	72	74
104	Asp	100	Arg	83	81	75
105	Ser	101	Ser	92	86	84
106	Leu	102	Ala	93	77	89
107	Asn	103	Ile	90	80	94
108	Asn	104	Asp	72	59	68
109	Arg	105	Ser	90	48	86
110	Met	106	Leu	92	21	97
111	Leu	107	Asn	64	9	75
112	Ser	108	Asn	58	4	51
113	Gln	109	Arg	96	31	83
114	Ile	110	Met	96	52	97
115	Trp	111	Leu	66	89	79
116	Ser	112	Ser	90	60	87
α-helix 7						
121	Leu	117	His	59	0	83
α-helix 8						
128	Ala	124	Pro	0	0	0
129	Gln	125	Ser	24	42	90
130	Leu	126	Cys	76	89	99
131	Asp	127	Ala	75	95	77
132	Arg	128	Ala	68	91	89
133	Ala	129	Gln	77	95	94
134	Lys	130	Leu	91	94	95
135	Arg	131	Asp	72	90	80
136	Asp	132	Arg	58	71	81
137	Ile	133	Ala	67	54	89
138	Val	134	Lys	84	88	96
139	Arg	135	Arg	38	78	79

Table 4 continued

Donor		Acceptor		45A4	53A6	54A7
140	Ser	136	Asp	40	64	90
141	Arg	137	Ile	73	87	89
α-helix 9						
148	Gln	144	Asp	97	96	98
149	Arg	145	Ser	87	49	92
150	Ala	146	Leu	84	51	96
151	Leu	147	Tyr	92	53	98
152	Thr	148	Gln	54	75	83
153	Thr	149	Arg	26	50	74
154	Ala	150	Ala	82	47	96

force fields conserved the elements of β-sheet. In the 45A4 and 53A6 simulations the β-sheet was flexible and partly lost during simulation. Despite a failure to maintain the helices, the NOE bound violations (Fig. 7) were significantly lower when using the Liu force field. This suggests that, in this case, there is only weak correlation between the instability of the elements of secondary structure and NOE violations. This has previously been observed for the protein HEWL (Eichenberger et al. 2010). The NOE violations can also be displayed as a time series (Fig. 8), in which case the distribution of the running average of the NOE violations is shown as a time series. After a few nanoseconds, this distribution does not change much.

CM consists primarily of α-helix (Fig. 1) and was most stable when simulated using the 54A7 force field (Fig. 9; Table 4). Using the 45A4 force field, most of the helices were maintained, although there was some formation of sheet structure. The 53A6 force field resulted in significant loss of structure. The large increase in RMSD after 18 ns in the 54A7 simulation is related to the fact that the secondary structure in the centre of the second helix was lost and a kink is formed. This small structural rearrangement, associated with rotation of a single torsional angle, had a large effect on the RMSD. In all the simulations, α-helices 1, 3, 5 and 8 were slightly shorter than suggested in the PDB entry 2FP2 (Table 4).

Table 5 Occurrence (%) of backbone hydrogen bonds of GCN occurring for at least 10% in one of the two simulations

Donor		Acceptor		53A6	54A7
4	Leu	2	Tyr	0	71
7	Glu	3	His	11	91
8	Val	4	Leu	40	83
9	Ala	4	Leu	19	0
9	Ala	5	Glu	31	76
10	Arg	5	Glu	35	12
10	Arg	6	Asn	35	75
11	Leu	6	Asn	51	22
11	Leu	7	Glu	36	59
12	Lys	7	Glu	12	38
12	Lys	8	Val	18	52
13	Lys	7	Glu	25	0
13	Lys	8	Val	0	35
13	Lys	9	Ala	13	45
14	Leu	7	Glu	27	0
14	Leu	9	Ala	0	19
14	Leu	10	Arg	10	55
15	Val	11	Leu	0	43
16	Gly	12	Lys	0	30

The secondary structure and backbone hydrogen bonds of the GCN peptide are shown in Fig. 10 and Table 5. The α-helical backbone hydrogen bonds were maintained

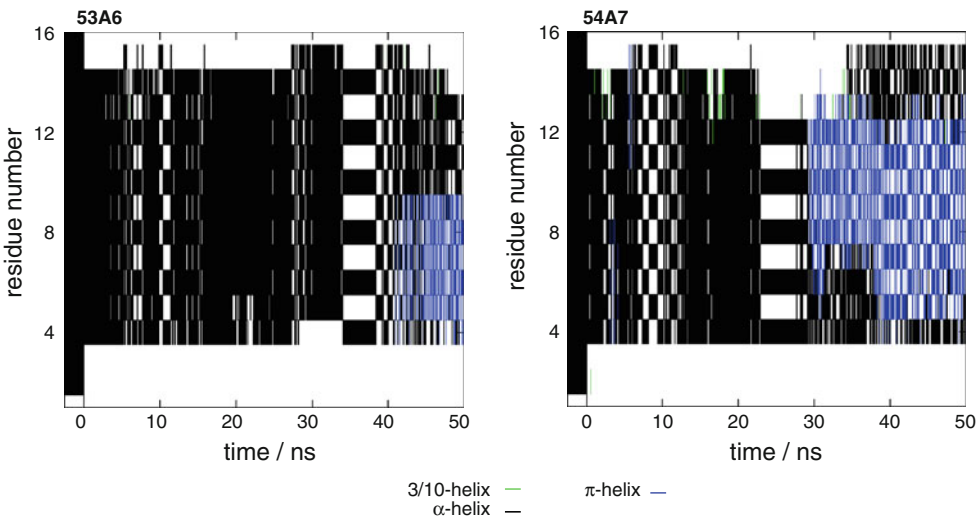

Fig. 10 Secondary structure of GCN in MD simulations using two different force fields. The definition of Kabsch and Sander (1983) is used. The secondary structure suggested in the PDB is shown in the box at the residue axis. Colour code: α-helix (*black*), 3_{10}-helix (*green*), π-helix (*blue*) and β-sheet (*red*)

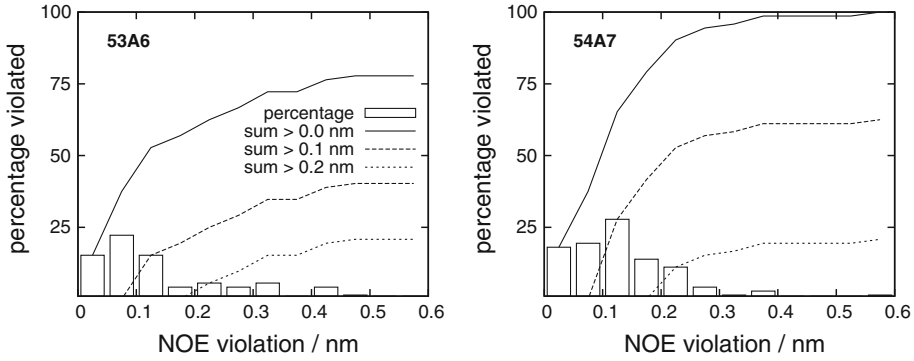

Fig. 11 NOE upper bound violations of GCN in MD simulations using four different force fields. Total number of NOE bounds: 72

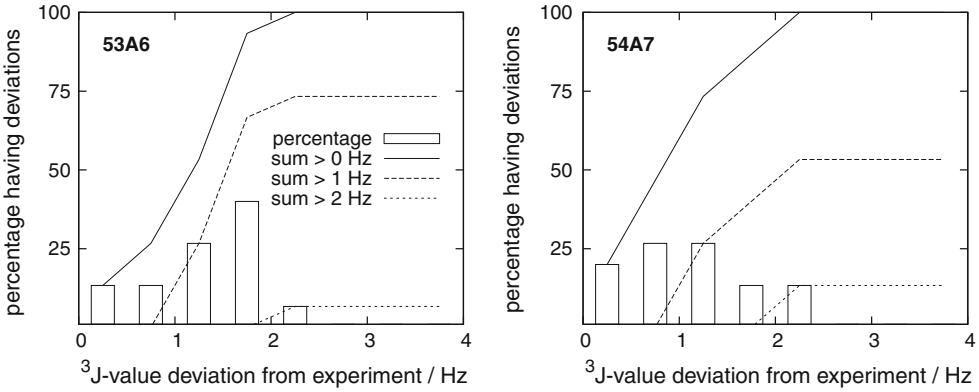

Fig. 12 ^3J-coupling deviations from measured values of GCN in MD simulations using four different force fields. Total number of ^3J-couplings: 15

during large parts of the 53A6 simulation, but after 40 ns a π-helix was partly formed. Using the 54A7 force field the π-helix was formed earlier after 30 ns and was also more pronounced. Comparing the measured NOE intensities with average distances from the simulation, it can be seen that the NOEs (Fig. 11) were less violated in the 53A6 simulation than in the 54A7 simulation. Although the total number of violations was significantly larger in the 54A7 simulation, the number of large violations (>0.2 nm) was similar in both runs. The ^3J-couplings (Fig. 12) showed similar deviations from the measured values in both simulations. For the GCN system, the 54A7 parameter set did not improve the agreement with the available experimental data. However, it is uncertain in this case whether this reflects a deficiency in the force field or the quality of the NMR data for GCN (Dolenc et al. 2010). We note that, also in this case, the NOE data are not of sufficient accuracy and quality to distinguish between a π- and an α-helical structure.

Conclusions

The ability of a new biomolecular force-field parameter set to improve the simulation results depends on the biomolecular system and on the properties or quantities considered. The 45A4 parameter set was best in retaining structure. The 53A6 parameter set improved the agreement with experimental data but had structural deficiencies in protein simulations. For the proteins considered, the Liu and 54A7 force fields generally improved the stability of secondary structure elements, while retaining the agreement with primary, observed data such as NOE intensities and ^3J-couplings. Because the 54A7 modification is simpler than the Liu modification of the 53A6 force field, the former was taken as the 54A7 GROMOS force field.

Acknowledgments We thank Hao Fan, Philippe Hünenberger, Zuo Le, Haiyan Liu, Alpesh Malde, Chris Oostenbrink, Xavier Perole, David Poger, Maria Reif, Denise Steiner, Xue Ying and Bojan Zagrovic for stimulating discussions and contributions to the force field modifications. This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology and by grant number 200020-121913 of the Swiss National Science Foundation, by grant number 228076 of the European Research Council and by grant number DP0770375 of the Australian Research Council. All funding is gratefully acknowledged.

References

Artymiuk PJ, Blake CCF, Rice DW, Wilson KS (1982) The structures of the monoclinic and orthorhombic forms of hen egg-white lysozyme at 6 angstroms resolution. Acta Crystallogr Sect B 38:778–783

Auweter SD, Fasan R, Reymond L, Underwood JG, Black DL, Pitsch S, Allain FH (2006) Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. Embo J 25:163–173

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullman B et al (eds) Intermolecular forces. Reidel, Dordrecht, The Netherlands, pp 331–342

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) Charmm—a program for macromolecular energy, minimization and dynamics calculations. J Comput Chem 4:187–217

Cao Z, Lin Z, Wang J, Liu H (2009) Refining the description of peptide backbone conformations improves protein simulations using the GROMOS 53A6 force field. J Comput Chem 30:645–660

Christen M, Hünenberger PH, Bakowies D, Baron R, Bürgi R, Geerke DP, Heinz TN, Kastenholz MA, Kräutler V, Oostenbrink C, Peter C, Trzesniak D, van Gunsteren WF (2005) The GROMOS software for biomolecular simulation: GROMOS05. J Comput Chem 26:1719–1751

Cornell WD, Cieplak P, Bayly CI, Gould IR, Mertz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 117:5179–5197

Daura X, Mark AE, van Gunsteren WF (1998) Parametrization of aliphatic CHn united atoms of GROMOS96 force field. J Comput Chem 19:535–547

Dolenc J, Missimer JH, Steinmetz MO, van Gunsteren WF (2010) Methods of NMR structure refinement: molecular dynamics simulations improve the agreement with measured NMR data of a C-terminal peptide of GCN4-p1. J Biomol NMR 47:221–235

Eichenberger AP, Gattin Z, Yalak G, van Gunsteren WF (2010) Molecular dynamics simulation of ester-linked hen egg white lysozyme reveals the effect of missing backbone hydrogen-bond donors on the protein structure. Helv Chim Acta 93:1857–1869

Fletcher CM, Jones DN, Diamond R, Neuhaus D (1996) Treatment of NOE constraints involving equivalent or nonstereoassigned protons in calculations of biomacromolecular structures. J Biomol NMR 169:292–310

Glättli A, van Gunsteren WF (2004) Are NMR-derived model structures for peptides representative for the ensemble of structures adopted in solution? Probing the fourth helical secondary structure of beta-peptides by molecular dynamics simulation. Angew Chem 116:6472–6476

Heinz TN, van Gunsteren WF, Huenenberger PH (2001) Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. J Chem Phys 115: 1125–1135

Humphrey W, Dalke A, Schulten K (1996) VMD—visual molecular dynamics. J Mol Graph 14:33–38

Jorgensen WL, Tirado-Rives J (1988) The OPLS potential functions for proteins—energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110:1657–1666

Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Karplus M (1959) Interpretation of the electron-spin resonance spectrum of the methyl radical. J. Chem. Phys. 30:11–15

MacKerell AD Jr, Wiórkiewicz-Kuczera J, Karplus M (1995) An all-atom empirical energy function for the simulation of nucleic-acids. J Am Chem Soc 117:11946–11975

MacKerell AD Jr, Bashford D, Bellot M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616

Okvist M, Dey R, Sasso S, Grahn E, Kast P, Krengel U (2006) 1.6Å crystal structure of the secreted chorismate mutase from mycobacterium tuberculosis. Novel Fold Topol Reveal 357:1483–1499

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656

Oostenbrink C, Soares TA, van der Vegt NFA, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. Eur Biophys J 34:273–284

Pardi A, Billeter M, Wüthrich K (1984) Calibration of the angular dependence of the amide proton-c alpha proton coupling constants, $^3J_{HN-\alpha}$, in a globular protein. use of $^3J_{HN-\alpha}$ for identification of helical secondary structure. J Mol Biol 180:741–751

Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman PA (1995) Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, moleculardynamics and free-energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 91:1–41

Poger D, van Gunsteren WF, Mark AE (2010) A new force field for simulating phosphatidylcholine bilayers. J Comput Chem 31:1117–1125

Reif M, Hünenberger PH (2010) Computation of methodology-independent single-ion solvation properties from molecular simulations. IV. Optimised Lennard–Jones interaction parameter sets for the alkali and halide ions in water. J Chem Phys 134:144104–144125

Ryckaert J-P, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23:327–341

Schuler LD, Daura X, van Gunsteren WF (2001) An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. J Comput Chem 22:1205–1218

Schwalbe H, Grimshaw SB, Spencer A, Buck M, Boyd J, Dobson CM, Redfield C, Smith LJ (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. Protein Sci 10:677–688

Steinmetz MO, Jelesarov I, Matousek WM, Honnappa S, Jahnke W, Missimer JH, Frank S, Alexandrescu AT, Kammerer RA (2007) Molecular basis of coiled-coil formation. PNAS 17:7062–7067

van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP, Tironi IG (1996) Biomolecular simulation: the GROMOS96 manual and user guide. Hochschulverlag AG, ETH Zurich

Weiner PK, Kollman PA (1981) Amber—assisted model-building with energy refinement—a general program for modeling molecules and their interactions. J Comput Chem 2:287–303

Wüthrich K, Billeter M, Braun W (1983) Pseudo-structures for the 20 common amino acids for use in studies of protein conformations by measurements of intramolecular proton-proton distance constraints with nuclear magnetic resonance. J Mol Biol 169:949–61